STAT 713 Feb 26 2019

Approach 1. Find CRLB $\dfrac{[\tau'(\theta)]^2}{-E\left[\frac{\partial^2}{\partial\theta^2}\ln f_{\underset{\sim}{X}}(\underset{\sim}{x}|\theta)\right]}$

↑ meets

if iid → $-E\left[\frac{\partial^2}{\partial\theta^2}\ln f_{\underset{\sim}{X}}(\underset{\sim}{x}|\theta)\right]$ ↓ $-n \times E\left[\frac{\partial^2}{\partial\theta^2}\ln f_X(x|\theta)\right]$

Find $\hat{\theta}$ which is unbiased

Attainment Corollary

if $S(\theta|\underset{\sim}{X}) = a(\theta)\left[W(\underset{\sim}{X}) - \tau(\theta)\right]$ and $W(\underset{\sim}{X})$ is an unbiased estimator for $\tau(\theta)$

then, $Var\left[W(\underset{\sim}{X})\right]$ attains CRLB (for $\tau(\theta)$) ; i.e., $W(\underset{\sim}{X})$ is the UMVUE of $\tau(\theta)$

Approach 2: based on sufficiency and completeness.

- If T is a sufficient and complete statistic for θ
  and there exists a function ϕ
  such that $E_\theta[\phi(T)] = \tau(\theta) \quad \forall \theta$

  Then $\phi(T)$ is the UMVUE of $\tau(\theta)$

why sufficiency?

Theorem 7.3.17 (Rao-Blackwell)

Let $W(\underset{\sim}{X})$ be an unbiased est. of $\tau(\theta)$

Let T be a "sufficient" statistic

$$\phi(T) = E[W \mid T]$$

Then
- $\phi(T)$ is also an unbiased est. of $\tau(\theta)$
- $Var[\phi(T)] \le Var[W(\underset{\sim}{X})]$

also implies: If UMVUE exists it must be a function of some sufficient statistic

why completeness.

① UMVUE is unique.a.s. (if exists)

② $W(\underset{\sim}{X})$ is the UMVUE of $\tau(\theta)$ if and only if $cov(W(\underset{\sim}{X}), U) = 0$ for any U where U is an unbiased estimator of 0.

$E_\theta(U) = 0$ for all θ.

proof of ①: suppose both W and W' are the UMVUE.

Let $W^* = \frac{1}{2}(W + W')$

we have $E[W^*] = \tau(\theta)$ for all θ.

Then W^* is an unbiased est. of $\tau(\theta)$

$$\begin{aligned}
Var(W^*) &= \frac{1}{4} Var(W) + \frac{1}{4} Var(W') + \frac{1}{2} Cov(W, W') \\
&= \frac{1}{2} Var(W) + \frac{1}{2} Cov(W, W') \\
&\le \frac{1}{2} Var(W) + \frac{1}{2} \sqrt{Var(W)\, Var(W')} \quad \text{(Cauchy-Swarchz Inequality)} \\
&= \frac{1}{2} Var(W) + \frac{1}{2} Var(W) = Var(W)
\end{aligned}$$

Because W is the UMVUE, $Var(W^*) \ge Var(W)$

$\Rightarrow Var(W^*) = Var(W) \Rightarrow Cov(W, W') = Var(W)$

$$Var(W-W') = Var(W) + Var(W') - 2Cov(W, W')$$

$$= Var(W) + Var(W') - 2Var(W)$$

$$= 0$$

Then $W-W'$ must be a constant (a.s.)

If $W-W' = c \neq 0$

$C = E[W-W'] = \tau(\theta) - \tau(\theta)$ contradiction

So $W-W' = 0$ (a.s.)

$W = W'$ (a.s.)

Proof of ② "$\Leftarrow\Rightarrow$"

start with "$\Leftarrow$" : meaning we want to prove that

If $Cov(W, U) = 0$ for all unbiased estimator U of 0

then W is the UMVUE

Suppose W' is ano. unbiased estimate of $\tau(\theta)$

$E[W'] = \tau(\theta)$ for θ

$E[W'-W] = 0$ for θ

$U = W'-W$ is an unbiased estimator of 0

$Cov(W, U) = Cov(W, W'-W) = 0$

$$Var(W') = Var(W + W'-W)$$

$$= Var(W) + Var(W'-W) + \underbrace{2Cov(W, W'-W)}_{0}$$

$$= Var(W) + Var(W'-W) \geq Var(W) \quad \checkmark$$

So we proved any unbiased est. of $\tau(\theta)$ has a no less variance that W, makin W the UMVUE

" $\Rightarrow$ " : means we need to prove

" if W is the UMVUE, then

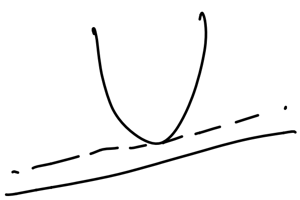
$Cov(W, U) = 0$ for all unbiased estimator U of 0. "

we have $W_1 = W + aU$ as unbiased esti of $\tau(\theta)$

Because W is the UMVUE

$Var(W) \leq var(W_1) = Var(W + aU) = Var(W) + a^2 Var(U) + 2a\, cov(W, U)$

$a^2 Var(U) + 2a\, cov(W, U) \geq 0$ hold for any $a \in R$



. it implies : $\Delta = [2\, cov(W, U)]^2 - 0 \leq 0$

$\Rightarrow cov(W, U) = 0$

Back to Approach 2, we now prove it.

T is a sufficient and complete statistic

and ϕ is a function such that $E[\phi(T)] = \tau(\theta)$ for all θ

Then $\phi(T)$ is the UMVUE of $\tau(\theta)$

proof: it suffices to show that

" for any unbiased estimator, U, of 0 ; i.e. $E[U] = 0$ for all θ

$cov[\phi(T), U] = 0$ "

This is true because. $Cov[\phi(T), U]$

$= E[\phi(T) U] - E[\phi(T)] E[U]$

$$= E[\phi(T)U] - E[\phi(T)] \times 0$$

$$= E[\phi(T)U]$$

"iterated" $= E[\,E[\phi(T)U \mid T]\,]$

$$= E[\,\phi(T)\,E[U \mid T]\,]$$

because T is sufficient, $E[U \mid T]$ does not depend on θ

Let $g(T) = E[U \mid T]$

$E[g(T)] = E[\,E[U \mid T]\,] \overset{\text{"iterated"}}{=} E[U] = 0$

because T is complete,

$P_\theta[g(t) = 0] = 1$; i.e., g is almost surly zero

So $E[U \mid T] = 0$ (a.s)

Thus $Cov[\phi(T), U] = E[\phi(T)\,E[U \mid T]]$

$$= E[\phi(T) \times 0] = 0$$

Hence $\phi(T)$ is the UMVUE of $\tau(\theta)$

7.3.3 Sufficiency and completeness

Remark: We now move to our "second approach" on how to find UMVUEs. This approach involves sufficiency and completeness−two topics we discussed in the last chapter. We can also address the unresolved issues on the previous page.

Theorem 7.3.17 (Rao-Blackwell). Let $W = W(\mathbf{X})$ be an unbiased estimator of $\tau(\theta)$. Let $T = T(\mathbf{X})$ be a sufficient statistic for θ. Define

$$\phi(T) = E(W|T).$$

Then

1. $E_\theta[\phi(T)] = \tau(\theta)$ for all $\theta \in \Theta$
2. $\text{var}_\theta[\phi(T)] \leq \text{var}_\theta(W)$ for all $\theta \in \Theta$.

That is, $\phi(T) = E(W|T)$ is a uniformly better unbiased estimator than W.
Proof. This result follows from the iterated rules for means and variances. First,

$$E_\theta[\phi(T)] = E_\theta[E(W|T)] = E_\theta(W) = \tau(\theta).$$

Second,

$$\begin{aligned} \text{var}_\theta(W) &= E_\theta[\text{var}(W|T)] + \text{var}_\theta[E(W|T)] \\ &= E_\theta[\text{var}(W|T)] + \text{var}_\theta[\phi(T)] \\ &\geq \text{var}_\theta[\phi(T)], \end{aligned}$$

because $\text{var}(W|T) \geq 0$ (a.s.) and hence $E_\theta[\text{var}(W|T)] \geq 0$. □

Implication: We can always "improve" the unbiased estimator W by conditioning on a sufficient statistic.

Remark: To use the Rao-Blackwell Theorem, some students think they have to

1. Find an unbiased estimator W.
2. Find a sufficient statistic T.
3. Derive the conditional distribution $f_{W|T}(w|t)$.
4. Find the mean $E(W|T)$ of this conditional distribution.

This is not the case at all! Because $\phi(T) = E(W|T)$ is a function of the sufficient statistic T, the Rao-Blackwell result simply convinces us that in our search for the UMVUE, we can restrict attention to those estimators that are functions of a sufficient statistic.

Q: In the proof of the Rao-Blackwell Theorem, where did we use the fact that T was sufficient?
A: Nowhere. Thus, it would seem that conditioning on any statistic, sufficient or not, will result in an improvement over the unbiased W. However, there is a catch:

- If T is not sufficient, then there is no guarantee that $\phi(T) = E(W|T)$ will be an estimator; i.e., it could depend on θ. See Example 7.3.18 (CB, pp 343).

Remark: To understand how we can use the Rao-Blackwell result in our quest to find a UMVUE, we need two additional results. One deals with uniqueness; the other describes an interesting characterization of a UMVUE itself.

Theorem 7.3.19 (**Uniqueness**). If W is UMVUE for $\tau(\theta)$, then it is unique.
Proof. Suppose that W' is also UMVUE. It suffices to show that $W = W'$ with probability one. Define

$$W^* = \frac{1}{2}(W + W').$$

Note that

$$E_\theta(W^*) = \frac{1}{2}[E_\theta(W) + E_\theta(W')] = \tau(\theta), \quad \text{for all } \theta \in \Theta,$$

showing that W^* is an unbiased estimator of $\tau(\theta)$. The variance of W^* is

$$\begin{aligned}
\text{var}_\theta(W^*) &= \text{var}_\theta\left[\frac{1}{2}(W + W')\right] \\
&= \frac{1}{4}\text{var}_\theta(W) + \frac{1}{4}\text{var}_\theta(W') + \frac{1}{2}\text{cov}_\theta(W, W') \\
&\leq \frac{1}{4}\text{var}_\theta(W) + \frac{1}{4}\text{var}_\theta(W') + \frac{1}{2}\left[\text{var}_\theta(W)\text{var}_\theta(W')\right]^{1/2} \\
&= \text{var}_\theta(W),
\end{aligned}$$

where the inequality arises from the covariance inequality (CB, pp 188, application of Cauchy-Schwarz) and the final equality holds because both W and W' are UMVUE by assumption (so their variances must be equal). Therefore, we have shown that

1. W^* is unbiased for $\tau(\theta)$
2. $\text{var}_\theta(W^*) \leq \text{var}_\theta(W)$.

Because W is UMVUE (by assumption), the inequality in (2) can not be strict (or else it would contradict the fact that W is UMVUE). Therefore, it must be true that

$$\text{var}_\theta(W^*) = \text{var}_\theta(W).$$

This implies that the inequality above (arising from the covariance inequality) is an equality; therefore,

$$\text{cov}_\theta(W, W') = \left[\text{var}_\theta(W)\text{var}_\theta(W')\right]^{1/2}.$$

Therefore,

$$\text{corr}_\theta(W, W') = \pm 1 \implies W' = \underbrace{a(\theta)W + b(\theta)}_{\text{linear function of } W}, \quad \text{with probability 1},$$

by Theorem 4.5.7 (CB, pp 172), where $a(\theta)$ and $b(\theta)$ are constants. It therefore suffices to show that $a(\theta) = 1$ and $b(\theta) = 0$. Note that

$$\begin{aligned} \text{cov}_\theta(W, W') = \text{cov}_\theta[W, a(\theta)W + b(\theta)] &= a(\theta)\text{cov}_\theta(W, W) \\ &= a(\theta)\text{var}_\theta(W). \end{aligned}$$

However, we have previously shown that

$$\begin{aligned} \text{cov}_\theta(W, W') = [\text{var}_\theta(W)\text{var}_\theta(W')]^{1/2} &= [\text{var}_\theta(W)\text{var}_\theta(W)]^{1/2} \\ &= \text{var}_\theta(W). \end{aligned}$$

This implies $a(\theta) = 1$. Finally,

$$\begin{aligned} E_\theta(W') = E_\theta[a(\theta)W + b(\theta)] &= E_\theta[W + b(\theta)] \\ &= E_\theta(W) + b(\theta). \end{aligned}$$

Because both W and W' are unbiased, this implies $b(\theta) = 0$. □

Theorem 7.3.20. Suppose $E_\theta(W) = \tau(\theta)$ for all $\theta \in \Theta$. W is UMVUE of $\tau(\theta)$ if and only if W is uncorrelated with all unbiased estimators of 0.
Proof. Necessity ($\Longrightarrow$): Suppose $E_\theta(W) = \tau(\theta)$ for all $\theta \in \Theta$. Suppose W is UMVUE of $\tau(\theta)$. Suppose $E_\theta(U) = 0$ for all $\theta \in \Theta$. It suffices to show $\text{cov}_\theta(W, U) = 0$ for all $\theta \in \Theta$. Define

$$\phi_a = W + aU,$$

where a is a constant. It is easy to see that ϕ_a is an unbiased estimator of $\tau(\theta)$; for all $\theta \in \Theta$,

$$E_\theta(\phi_a) = E_\theta(W + aU) = E_\theta(W) + a\underbrace{E_\theta(U)}_{=\,0} = \tau(\theta).$$

Also,

$$\begin{aligned} \text{var}_\theta(\phi_a) &= \text{var}_\theta(W + aU) \\ &= \text{var}_\theta(W) + \underbrace{a^2\text{var}_\theta(U) + 2a\ \text{cov}_\theta(W, U)}_{\text{Key question: Can this be negative?}}. \end{aligned}$$

- **Case 1:** Suppose $\exists\ \theta_0 \in \Theta$ such that $\text{cov}_{\theta_0}(W, U) < 0$. Then

$$\begin{aligned} a^2\text{var}_{\theta_0}(U) + 2a\ \text{cov}_{\theta_0}(W, U) < 0 &\iff a^2\text{var}_{\theta_0}(U) < -2a\ \text{cov}_{\theta_0}(W, U) \\ &\iff a^2 < -\frac{2a\ \text{cov}_{\theta_0}(W, U)}{\text{var}_{\theta_0}(U)}. \end{aligned}$$

I can make this true by picking

$$0 < a < -\frac{2\ \text{cov}_{\theta_0}(W,U)}{\text{var}_{\theta_0}(U)}$$

and therefore I have shown that

$$\text{var}_{\theta_0}(\phi_a) < \text{var}_{\theta_0}(W).$$

However, this contradicts the assumption that W is UMVUE. Therefore, it must be true that $\text{cov}_\theta(W,U) \geq 0$.

- **Case 2:** Suppose $\exists\ \theta_0 \in \Theta$ such that $\text{cov}_{\theta_0}(W,U) > 0$. Then

$$\begin{aligned} a^2\text{var}_{\theta_0}(U) + 2a\ \text{cov}_{\theta_0}(W,U) < 0 &\iff a^2\text{var}_{\theta_0}(U) < -2a\ \text{cov}_{\theta_0}(W,U) \\ &\iff a^2 < -\frac{2a\ \text{cov}_{\theta_0}(W,U)}{\text{var}_{\theta_0}(U)}. \end{aligned}$$

I can make this true by picking

$$-\frac{2\ \text{cov}_{\theta_0}(W,U)}{\text{var}_{\theta_0}(U)} < a < 0$$

and therefore I have shown that

$$\text{var}_{\theta_0}(\phi_a) < \text{var}_{\theta_0}(W).$$

However, this again contradicts the assumption that W is UMVUE. Therefore, it must be true that $\text{cov}_\theta(W,U) \leq 0$.

Combining Case 1 and Case 2, we are forced to conclude that $\text{cov}_\theta(W,U) = 0$. This proves the necessity.

Sufficiency ($\Longleftarrow$): Suppose $E_\theta(W) = \tau(\theta)$ for all $\theta \in \Theta$. Suppose $\text{cov}_\theta(W,U) = 0$ for all $\theta \in \Theta$ where U is any unbiased estimator of zero; i.e., $E_\theta(U) = 0$ for all $\theta \in \Theta$. Let W' be any other unbiased estimator of $\tau(\theta)$. It suffices to show that $\text{var}_\theta(W) \leq \text{var}_\theta(W')$. Write

$$W' = W + (W' - W)$$

and calculate

$$\text{var}_\theta(W') = \text{var}_\theta(W) + \text{var}_\theta(W' - W) + 2\text{cov}_\theta(W, W' - W).$$

However, $\text{cov}_\theta(W, W' - W) = 0$ because $W' - W$ is an unbiased estimator of 0. Therefore,

$$\text{var}_\theta(W') = \text{var}_\theta(W) + \underbrace{\text{var}_\theta(W' - W)}_{\geq 0} \ \geq\ \text{var}_\theta(W).$$

This proves the sufficiency. □

Summary: We are now ready to put Theorem 7.3.17 (Rao-Blackwell), Theorem 7.3.19 (UMVUE uniqueness) and Theorem 7.3.20 together. Suppose $\mathbf{X} \sim f_{\mathbf{X}}(\mathbf{x}|\theta)$, where $\theta \in \Theta$. Our goal is to find the UMVUE of $\tau(\theta)$.

- Theorem 7.3.17 (Rao-Blackwell) assures us that we can restrict attention to functions of sufficient statistics.

Therefore, suppose T is a sufficient statistic for θ. Suppose that $\phi(T)$, a function of T, is an unbiased estimator of $\tau(\theta)$; i.e.,

$$E_\theta[\phi(T)] = \tau(\theta), \quad \text{for all } \theta \in \Theta.$$

- Theorem 7.3.20 assures us that $\phi(T)$ is UMVUE if and only if $\phi(T)$ is uncorrelated with all unbiased estimators of 0.

Add the assumption that T is a complete statistic. *The only unbiased estimator of 0 in complete families is the zero function itself.* Because $\text{cov}_\theta[\phi(T), 0] = 0$ holds trivially, we have shown that $\phi(T)$ is uncorrelated with "all" unbiased estimators of 0. Theorem 7.3.20 says that $\phi(T)$ must be UMVUE; Theorem 7.3.19 guarantees that $\phi(T)$ is unique.

Recipe for finding UMVUEs: Suppose we want to find the UMVUE for $\tau(\theta)$.

1. Start by finding a statistic T that is both sufficient and complete.
2. Find a function of T, say $\phi(T)$, that satisfies

$$E_\theta[\phi(T)] = \tau(\theta), \quad \text{for all } \theta \in \Theta.$$

Then $\phi(T)$ is the UMVUE for $\tau(\theta)$. This is essentially what is summarized in Theorem 7.3.23 (CB, pp 347).

Example 7.17. Suppose $X_1, X_2, ..., X_n$ are iid Poisson(θ), where $\theta > 0$.

- We already know that $\overline{X}$ is UMVUE for θ; we proved this by showing that $\overline{X}$ is unbiased and that $\text{var}_\theta(\overline{X})$ attains the CRLB on the variance of all unbiased estimators of θ.
- We now show $\overline{X}$ is UMVUE for θ by using sufficiency and completeness.

The pmf of X is

$$\begin{aligned} f_X(x|\theta) &= \frac{\theta^x e^{-\theta}}{x!} I(x = 0, 1, 2, ...,) \\ &= \frac{I(x = 0, 1, 2, ...,)}{x!} e^{-\theta} e^{(\ln\theta)x} \\ &= h(x)c(\theta)\exp\{w_1(\theta)t_1(x)\}. \end{aligned}$$

Therefore X has pmf in the exponential family. Theorem 6.2.10 says that

$$T = T(\mathbf{X}) = \sum_{i=1}^{n} X_i$$

is a sufficient statistic. Because $d = k = 1$ (i.e., a full family), Theorem 6.2.25 says that T is complete. Now,

$$E_\theta(T) = E_\theta\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E_\theta(X_i) = n\theta.$$

Therefore,

$$E_\theta\left(\frac{T}{n}\right) = E_\theta(\overline{X}) = \theta.$$

Because $\overline{X}$ is unbiased and is a function of T, a complete and sufficient statistic, we know that $\overline{X}$ is the UMVUE.

Example 7.18. Suppose $X_1, X_2, ..., X_n$ are iid $\mathcal{U}(0, \theta)$, where $\theta > 0$. We have previously shown that

$$T = T(\mathbf{X}) = X_{(n)}$$

is sufficient and complete (see Example 6.5 and Example 6.16, respectively, in the notes). It follows that

$$E_\theta(T) = E_\theta(X_{(n)}) = \left(\frac{n}{n+1}\right)\theta$$

for all $\theta > 0$. Therefore,

$$E_\theta\left[\left(\frac{n+1}{n}\right) X_{(n)}\right] = \theta.$$

Because $(n+1)X_{(n)}/n$ is unbiased and is a function of $X_{(n)}$, a complete and sufficient statistic, it must be the UMVUE.

Example 7.19. Suppose $X_1, X_2, ..., X_n$ are iid gamma(α_0, β), where α_0 is known and $\beta > 0$. Find the UMVUE of $\tau(\beta) = 1/\beta$.
Solution. The pdf of X is

$$\begin{aligned} f_X(x|\beta) &= \frac{1}{\Gamma(\alpha_0)\beta^{\alpha_0}} x^{\alpha_0 - 1} e^{-x/\beta} I(x > 0) \\ &= \frac{x^{\alpha_0 - 1} I(x > 0)}{\Gamma(\alpha_0)} \, \frac{1}{\beta^{\alpha_0}} \, e^{(-1/\beta)x} \\ &= h(x)c(\beta)\exp\{w_1(\beta)t_1(x)\} \end{aligned}$$

a one-parameter exponential family with $d = k = 1$ (a full family). Theorem 6.2.10 and Theorem 6.2.25 assure that

$$T = T(\mathbf{X}) = \sum_{i=1}^{n} X_i$$

is a sufficient and complete statistic, respectively. In Example 7.16 (notes), we saw that

$$\phi(T) = \frac{n\alpha_0 - 1}{T}$$

is an unbiased estimator of $\tau(\beta) = 1/\beta$. Therefore, $\phi(T)$ must be the UMVUE.

Remark: In Example 7.16, recall that the CRLB on the variance of unbiased estimators of $\tau(\beta) = 1/\beta$ was unattainable.

Example 7.20. Suppose $X_1, X_2, ..., X_n$ are iid Poisson(θ), where $\theta > 0$. Find the UMVUE for

$$\tau(\theta) = P_\theta(X = 0) = e^{-\theta}.$$

Solution. We use an approach known as "direct conditioning." We start with

$$T = T(\mathbf{X}) = \sum_{i=1}^{n} X_i,$$

which is sufficient and complete. We know that the UMVUE therefore is a function of T. Consider forming

$$\phi(T) = E(W|T),$$

where W is any unbiased estimator of $\tau(\theta) = e^{-\theta}$. We know that $\phi(T)$ by this construction is the UMVUE; clearly $\phi(T) = E(W|T)$ is a function of T and

$$E_\theta[\phi(T)] = E_\theta[E(W|T)] = E_\theta(W) = e^{-\theta}.$$

How should we choose W? Any unbiased W will "work," so let's keep our choice simple, say

$$W = W(\mathbf{X}) = I(X_1 = 0).$$

Note that

$$E_\theta(W) = E_\theta[I(X_1 = 0)] = P_\theta(X_1 = 0) = e^{-\theta},$$

showing that W is an unbiased estimator. Now, we just calculate $\phi(T) = E(W|T)$ directly. For t fixed, we have

$$\begin{aligned}
\phi(t) = E(W|T = t) &= E[I(X_1 = 0)|T = t] \\
&= P(X_1 = 0|T = t) \\
&= \frac{P_\theta(X_1 = 0, T = t)}{P_\theta(T = t)} \\
&= \frac{P_\theta\left(X_1 = 0, \sum_{i=2}^{n} X_i = t\right)}{P_\theta(T = t)} \\
&\overset{\text{indep}}{=} \frac{P_\theta(X_1 = 0)P_\theta\left(\sum_{i=2}^{n} X_i = t\right)}{P_\theta(T = t)}.
\end{aligned}$$

We can now calculate each of these probabilities. Recall that $X_1 \sim \text{Poisson}(\theta)$, $\sum_{i=2}^n X_i \sim \text{Poisson}((n-1)\theta)$, and $T \sim \text{Poisson}(n\theta)$. Therefore,

$$\begin{aligned} \phi(t) &= \frac{P_\theta(X_1 = 0) P_\theta\left(\sum_{i=2}^n X_i = t\right)}{P_\theta(T = t)} \\ &= \frac{e^{-\theta}\, \dfrac{[(n-1)\theta]^t e^{-(n-1)\theta}}{t!}}{\dfrac{(n\theta)^t e^{-n\theta}}{t!}} = \left(\frac{n-1}{n}\right)^t. \end{aligned}$$

Therefore,

$$\phi(T) = \left(\frac{n-1}{n}\right)^T$$

is the UMVUE of $\tau(\theta) = e^{-\theta}$.

Remark: It is interesting to note that in this example

$$\phi(t) = \left(\frac{n-1}{n}\right)^t = \left[\left(\frac{n-1}{n}\right)^n\right]^{\overline{x}} = \left[\left(1 - \frac{1}{n}\right)^n\right]^{\overline{x}} \approx e^{-\overline{x}},$$

for n large. Recall that $e^{-\overline{X}}$ is the MLE of $\tau(\theta) = e^{-\theta}$ by invariance.

Remark: The last subsection in CB (Section 7.3.4) is on loss-function optimality. This material will be covered in STAT 822.

7.4 Appendix: CRLB Theory

Remark: In this section, we provide the proofs that pertain to the CRLB approach to finding UMVUEs. These proofs are also relevant for later discussions on MLEs and their large-sample characteristics.

Remark: We start by reviewing the Cauchy-Schwarz Inequality. Essentially, the main Cramér-Rao inequality result (Theorem 7.3.9) follows as an application of this inequality.

Recall: Suppose X and Y are random variables. Then

$$|E(XY)| \leq E(|XY|) \leq [E(X^2)]^{1/2}[E(Y^2)]^{1/2}.$$

This is called the **Cauchy-Schwarz Inequality**. In this inequality, if we replace X with $X - \mu_X$ and Y with $Y - \mu_Y$, we get

$$|E[(X-\mu_X)(Y-\mu_Y)]| \leq \{E[(X-\mu_X)^2]\}^{1/2}\{E[(Y-\mu_Y)^2]\}^{1/2}.$$

Squaring both sides, we get

$$[\text{cov}(X,Y)]^2 \leq \sigma_X^2 \sigma_Y^2.$$

This is called the **covariance inequality**.

Theorem 7.3.9 (Cramér-Rao Inequality). Suppose $\mathbf{X} \sim f_{\mathbf{X}}(\mathbf{x}|\theta)$, where

1. the support of $\mathbf{X}$ is free of all unknown parameters
2. for any function $h(\mathbf{x})$ such that $E_\theta[h(\mathbf{X})] < \infty$ for all $\theta \in \Theta$, the interchange

$$\frac{d}{d\theta}\int_{\mathbb{R}^n} h(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}|\theta) d\mathbf{x} = \int_{\mathbb{R}^n} \frac{\partial}{\partial\theta} h(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}|\theta) d\mathbf{x}$$

is justified; i.e., we can interchange the derivative and integral (derivative and sum if $\mathbf{X}$ is discrete).

For any estimator $W(\mathbf{X})$ with $\text{var}_\theta[W(\mathbf{X})] < \infty$, the following inequality holds:

$$\text{var}_\theta[W(\mathbf{X})] \ \geq \ \frac{\left\{\frac{d}{d\theta} E_\theta[W(\mathbf{X})]\right\}^2}{E_\theta\left\{\left[\frac{\partial}{\partial\theta} \ln f_{\mathbf{X}}(\mathbf{X}|\theta)\right]^2\right\}}.$$

Proof. First we state and prove a lemma.

Lemma. Let

$$S(\theta|\mathbf{X}) = \frac{\partial}{\partial\theta} \ln f_{\mathbf{X}}(\mathbf{X}|\theta)$$

denote the **score function**. The score function is a zero-mean random variable; that is,

$$E_\theta[S(\theta|\mathbf{X})] = E_\theta\left[\frac{\partial}{\partial\theta} \ln f_{\mathbf{X}}(\mathbf{X}|\theta)\right] = 0.$$

Proof of Lemma: Note that

$$\begin{aligned} E_\theta\left[\frac{\partial}{\partial\theta} \ln f_{\mathbf{X}}(\mathbf{X}|\theta)\right] = \int_{\mathbb{R}^n} \frac{\partial}{\partial\theta} \ln f_{\mathbf{X}}(\mathbf{x}|\theta) f_{\mathbf{X}}(\mathbf{x}|\theta) d\mathbf{x} &= \int_{\mathbb{R}^n} \frac{\frac{\partial}{\partial\theta} f_{\mathbf{X}}(\mathbf{x}|\theta)}{f_{\mathbf{X}}(\mathbf{x}|\theta)} f_{\mathbf{X}}(\mathbf{x}|\theta) d\mathbf{x} \\ &= \int_{\mathbb{R}^n} \frac{\partial}{\partial\theta} f_{\mathbf{X}}(\mathbf{x}|\theta) d\mathbf{x} \\ &= \frac{d}{d\theta} \underbrace{\int_{\mathbb{R}^n} f_{\mathbf{X}}(\mathbf{x}|\theta) d\mathbf{x}}_{=\,1} = 0. \end{aligned}$$

The interchange of derivative and integral above is justified based on the assumptions stated in Theorem 7.3.9. Therefore, the lemma is proven. □

Note: Because the score function is a zero-mean random variable,

$$\text{var}_\theta[S(\theta|\mathbf{X})] = E_\theta\{[S(\theta|\mathbf{X})]^2\};$$

that is,

$$\text{var}_\theta\left[\frac{\partial}{\partial\theta} \ln f_{\mathbf{X}}(\mathbf{X}|\theta)\right] = E_\theta\left\{\left[\frac{\partial}{\partial\theta} \ln f_{\mathbf{X}}(\mathbf{X}|\theta)\right]^2\right\}.$$

We now return to the CRLB proof. Consider

$$\begin{aligned}
\text{cov}_\theta\left[W(\mathbf{X}), \frac{\partial}{\partial\theta}\ln f_{\mathbf{X}}(\mathbf{X}|\theta)\right] &= E_\theta\left[W(\mathbf{X})\frac{\partial}{\partial\theta}\ln f_{\mathbf{X}}(\mathbf{X}|\theta)\right] - E_\theta[W(\mathbf{X})]\underbrace{E_\theta\left[\frac{\partial}{\partial\theta}\ln f_{\mathbf{X}}(\mathbf{X}|\theta)\right]}_{=\,0} \\
&= E_\theta\left[W(\mathbf{X})\frac{\partial}{\partial\theta}\ln f_{\mathbf{X}}(\mathbf{X}|\theta)\right] \\
&= \int_{\mathbb{R}^n} W(\mathbf{x})\frac{\partial}{\partial\theta}\ln f_{\mathbf{X}}(\mathbf{x}|\theta) f_{\mathbf{X}}(\mathbf{x}|\theta)d\mathbf{x} \\
&= \int_{\mathbb{R}^n} W(\mathbf{x})\frac{\frac{\partial}{\partial\theta}f_{\mathbf{X}}(\mathbf{x}|\theta)}{f_{\mathbf{X}}(\mathbf{x}|\theta)} f_{\mathbf{X}}(\mathbf{x}|\theta)d\mathbf{x} \\
&= \int_{\mathbb{R}^n} W(\mathbf{x})\frac{\partial}{\partial\theta}f_{\mathbf{X}}(\mathbf{x}|\theta)d\mathbf{x} \\
&= \frac{d}{d\theta}\int_{\mathbb{R}^n} W(\mathbf{x})f_{\mathbf{X}}(\mathbf{x}|\theta)d\mathbf{x} \\
&= \frac{d}{d\theta}E_\theta[W(\mathbf{X})].
\end{aligned}$$

Now, write the covariance inequality with

1. $W(\mathbf{X})$ playing the role of "X"
2. $S(\theta|\mathbf{X}) = \frac{\partial}{\partial\theta}\ln f_{\mathbf{X}}(\mathbf{X}|\theta)$ playing the role of "Y."

We get

$$\left\{\text{cov}_\theta\left[W(\mathbf{X}), \frac{\partial}{\partial\theta}\ln f_{\mathbf{X}}(\mathbf{X}|\theta)\right]\right\}^2 \leq \text{var}_\theta[W(\mathbf{X})]\ \text{var}_\theta\left[\frac{\partial}{\partial\theta}\ln f_{\mathbf{X}}(\mathbf{X}|\theta)\right],$$

that is,

$$\left\{\frac{d}{d\theta}E_\theta[W(\mathbf{X})]\right\}^2 \leq \text{var}_\theta[W(\mathbf{X})]\ E_\theta\left\{\left[\frac{\partial}{\partial\theta}\ln f_{\mathbf{X}}(\mathbf{X}|\theta)\right]^2\right\}.$$

Dividing both sides by $E_\theta\left\{\left[\frac{\partial}{\partial\theta}\ln f_{\mathbf{X}}(\mathbf{X}|\theta)\right]^2\right\}$ gives the result. □

Corollary 7.3.10 (Cramér-Rao Inequality−iid case). With the same regularity conditions stated in Theorem 7.3.9, in the iid case,

$$\text{var}_\theta[W(\mathbf{X})] \ \geq\ \frac{\left\{\frac{d}{d\theta}E_\theta[W(\mathbf{X})]\right\}^2}{nE_\theta\left\{\left[\frac{\partial}{\partial\theta}\ln f_X(X|\theta)\right]^2\right\}}.$$

Proof. It suffices to show

$$E_\theta\left\{\left[\frac{\partial}{\partial\theta}\ln f_{\mathbf{X}}(\mathbf{X}|\theta)\right]^2\right\} = nE_\theta\left\{\left[\frac{\partial}{\partial\theta}\ln f_X(X|\theta)\right]^2\right\}.$$

Because $X_1, X_2, ..., X_n$ are iid,

$$
\begin{aligned}
\text{LHS} &= E_\theta\left\{\left[\frac{\partial}{\partial\theta}\ln\prod_{i=1}^n f_X(X_i|\theta)\right]^2\right\} \\
&= E_\theta\left\{\left[\frac{\partial}{\partial\theta}\sum_{i=1}^n \ln f_X(X_i|\theta)\right]^2\right\} \\
&= E_\theta\left\{\left[\sum_{i=1}^n \frac{\partial}{\partial\theta}\ln f_X(X_i|\theta)\right]^2\right\} \\
&= \sum_{i=1}^n E_\theta\left\{\left[\frac{\partial}{\partial\theta}\ln f_X(X_i|\theta)\right]^2\right\} + \sum\sum_{i\neq j} E_\theta\left[\frac{\partial}{\partial\theta}\ln f_X(X_i|\theta)\frac{\partial}{\partial\theta}\ln f_X(X_j|\theta)\right] \\
&\overset{\text{indep}}{=} \sum_{i=1}^n E_\theta\left\{\left[\frac{\partial}{\partial\theta}\ln f_X(X_i|\theta)\right]^2\right\} + \sum\sum_{i\neq j} \underbrace{E_\theta\left[\frac{\partial}{\partial\theta}\ln f_X(X_i|\theta)\right]}_{=\,0}\underbrace{E_\theta\left[\frac{\partial}{\partial\theta}\ln f_X(X_j|\theta)\right]}_{=\,0}.
\end{aligned}
$$

Therefore, all cross product expectations are zero and thus

$$
\text{LHS} = \sum_{i=1}^n E_\theta\left\{\left[\frac{\partial}{\partial\theta}\ln f_X(X_i|\theta)\right]^2\right\} \overset{\text{ident}}{=} nE_\theta\left\{\left[\frac{\partial}{\partial\theta}\ln f_X(X|\theta)\right]^2\right\}.
$$

This proves the iid case. □

Remark: Recall our notation:

$$
\begin{aligned}
I_n(\theta) &= E_\theta\left\{\left[\frac{\partial}{\partial\theta}\ln f_{\mathbf{X}}(\mathbf{X}|\theta)\right]^2\right\} \\
I_1(\theta) &= E_\theta\left\{\left[\frac{\partial}{\partial\theta}\ln f_X(X|\theta)\right]^2\right\}.
\end{aligned}
$$

In the iid case, we have just proven that $I_n(\theta) = nI(\theta)$. Therefore, in the iid case,

- If $W(\mathbf{X})$ is an unbiased estimator of $\tau(\theta)$, then

$$
\text{CRLB} = \frac{[\tau'(\theta)]^2}{nI_1(\theta)}.
$$

- If $W(\mathbf{X})$ is an unbiased estimator of $\tau(\theta) = \theta$, then

$$
\text{CRLB} = \frac{1}{nI_1(\theta)}.
$$

Lemma 7.3.11 (Information Equality). Under regularity conditions,

$$I_1(\theta) = E_\theta\left\{\left[\frac{\partial}{\partial\theta}\ln f_X(X|\theta)\right]^2\right\} = -E_\theta\left[\frac{\partial^2}{\partial\theta^2}\ln f_X(X|\theta)\right].$$

Proof. From the definition of mathematical expectation,

$$E_\theta\left[\frac{\partial^2}{\partial\theta^2}\ln f_X(X|\theta)\right] = \int_{\mathbb{R}}\frac{\partial^2}{\partial\theta^2}\ln f_X(x|\theta) f_X(x|\theta)dx = \int_{\mathbb{R}}\underbrace{\frac{\partial}{\partial\theta}\left[\frac{\frac{\partial}{\partial\theta}f_X(x|\theta)}{f_X(x|\theta)}\right]}_{\text{use quotient rule here}} f_X(x|\theta)dx$$

Note: A sum replaces the integral above if X is discrete. The derivative

$$\begin{aligned}\frac{\partial}{\partial\theta}\left[\frac{\frac{\partial}{\partial\theta}f_X(x|\theta)}{f_X(x|\theta)}\right] &= \frac{\frac{\partial^2}{\partial\theta^2}f_X(x|\theta)f_X(x|\theta) - \frac{\partial}{\partial\theta}f_X(x|\theta)\frac{\partial}{\partial\theta}f_X(x|\theta)}{[f_X(x|\theta)]^2} \\ &= \frac{\frac{\partial^2}{\partial\theta^2}f_X(x|\theta)}{f_X(x|\theta)} - \frac{\left[\frac{\partial}{\partial\theta}f_X(x|\theta)\right]^2}{[f_X(x|\theta)]^2}.\end{aligned}$$

Therefore, the last integral becomes

$$\begin{aligned}\int_{\mathbb{R}}\left\{\frac{\frac{\partial^2}{\partial\theta^2}f_X(x|\theta)}{f_X(x|\theta)} - \frac{\left[\frac{\partial}{\partial\theta}f_X(x|\theta)\right]^2}{[f_X(x|\theta)]^2}\right\}f_X(x|\theta)dx &= \int_{\mathbb{R}}\left\{\frac{\partial^2}{\partial\theta^2}f_X(x|\theta) - \frac{\left[\frac{\partial}{\partial\theta}f_X(x|\theta)\right]^2}{f_X(x|\theta)}\right\}dx \\ &= \int_{\mathbb{R}}\frac{\partial^2}{\partial\theta^2}f_X(x|\theta)dx - \int_{\mathbb{R}}\frac{\left[\frac{\partial}{\partial\theta}f_X(x|\theta)\right]^2}{f_X(x|\theta)}\,dx \\ &= \frac{d^2}{d\theta^2}\underbrace{\int_{\mathbb{R}}f_X(x|\theta)dx}_{=1} - \int_{\mathbb{R}}\left[\frac{\partial}{\partial\theta}\ln f_X(x|\theta)\right]^2 f_X(x|\theta)dx \\ &= -E_\theta\left\{\left[\frac{\partial}{\partial\theta}\ln f_X(X|\theta)\right]^2\right\}.\end{aligned}$$

We have shown

$$E_\theta\left[\frac{\partial^2}{\partial\theta^2}\ln f_X(X|\theta)\right] = -E_\theta\left\{\left[\frac{\partial}{\partial\theta}\ln f_X(X|\theta)\right]^2\right\}.$$

Multiplying both sides by -1 gives the information equality. □

Remark: We now finish by proving the attainment result.

Corollary 7.3.15. Suppose $X_1, X_2, ..., X_n$ is an iid sample from $f_X(x|\theta)$, where $\theta \in \Theta$, a family that satisfies the regularity conditions stated for the Cramér-Rao Inequality. If $W(\mathbf{X})$ is an unbiased estimator of $\tau(\theta)$, then $\text{var}_\theta[W(\mathbf{X})]$ attains the CRLB if and only if the score function

$$S(\theta|\mathbf{x}) = a(\theta)[W(\mathbf{x}) - \tau(\theta)]$$

is a linear function of $W(\mathbf{x})$.

Proof. From the CRLB proof, recall that we had

1. $W(\mathbf{X})$ playing the role of "X"
2. $\frac{\partial}{\partial\theta}\ln f_{\mathbf{X}}(\mathbf{X}|\theta)$ playing the role of "Y"

in applying the covariance inequality, which yields

$$\begin{aligned}\text{var}_\theta[W(\mathbf{X})] &\geq \frac{[\tau'(\theta)]^2}{E_\theta\left\{\left[\frac{\partial}{\partial\theta}\ln f_{\mathbf{X}}(\mathbf{X}|\theta)\right]^2\right\}} \\ &\overset{\text{iid}}{=} \frac{[\tau'(\theta)]^2}{E_\theta\left\{\left[\frac{\partial}{\partial\theta}\ln\prod_{i=1}^n f_X(X_i|\theta)\right]^2\right\}}.\end{aligned}$$

Now, in the covariance inequality, we have *equality* when the correlation of $W(\mathbf{X})$ and $\frac{\partial}{\partial\theta}\ln f_{\mathbf{X}}(\mathbf{X}|\theta)$ equals ± 1, which in turn implies

$$c(X-\mu_X) = Y - \mu_Y \quad \text{a.s.},$$

or restated,

$$c[W(\mathbf{X})-\tau(\theta)] = \frac{\partial}{\partial\theta}\ln f_{\mathbf{X}}(\mathbf{X}|\theta) - 0 \quad \text{a.s.}$$

This is an application of Theorem 4.5.7 (CB, pp 172); i.e., two random variables are perfectly correlated if and only if the random variables are perfectly linearly related. In these equations, c is a constant. Also, I have written "-0" on the RHS of the last equation to emphasize that

$$E_\theta\left[\frac{\partial}{\partial\theta}\ln f_{\mathbf{X}}(\mathbf{X}|\theta)\right] = E_\theta\left[\frac{\partial}{\partial\theta}\ln\prod_{i=1}^n f_X(X_i|\theta)\right] = 0.$$

Also, $W(\mathbf{X})$ is an unbiased estimator of $\tau(\theta)$ by assumption. Therefore, we have

$$\begin{aligned}c[W(\mathbf{X})-\tau(\theta)] &= \frac{\partial}{\partial\theta}\ln f_{\mathbf{X}}(\mathbf{X}|\theta) \\ &= \frac{\partial}{\partial\theta}\ln\prod_{i=1}^n f_X(X_i|\theta) \\ &= \frac{\partial}{\partial\theta}\ln L(\theta|\mathbf{X}) \\ &= S(\theta|\mathbf{X}),\end{aligned}$$

where $S(\theta|\mathbf{X})$ is the score function. The constant c cannot depend on $W(\mathbf{X})$ nor on $\frac{\partial}{\partial\theta}\ln f_{\mathbf{X}}(\mathbf{X}|\theta)$, but it can depend on θ. To emphasize this, we write

$$S(\theta|\mathbf{X}) = a(\theta)[W(\mathbf{X})-\tau(\theta)].$$

Thus, $\text{var}_\theta[W(\mathbf{X})]$ attains the CRLB when the score function $S(\theta|\mathbf{X})$ can be written as a linear function of the unbiased estimator $W(\mathbf{X})$. □